DOCUMENT RESUME

ED 420 682 TM 028 357

AUTHOR Coleman, Donald G.; Areglado, Ron; Adams, R. C.
TITLE Establishing Construct Validity and Reliability for the NAESP Professional Development Inventory: Simplifying Assessment Center Techniques.
PUB DATE 1998-04-00
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Assessment Centers (Personnel); *Construct Validity; Elementary Education; Evaluation Methods; Factor Analysis; Instructional Leadership; Job Skills; *Principals; *Professional Development; *Reliability; Simulation
IDENTIFIERS *National Association Elementary School Principals

ABSTRACT

This report documents the construct validity and reliability of the Professional Development Inventory (PDI), an assessment center sponsored by the National Association of Elementary School Principals (NAESP). The assessment center provides a means for assessing participants in situations simulating those confronted by principals on the job. The NAESP has developed a new version of the assessment center, which focuses on 13 skills for principals. The time for assessing candidates has been dropped from 2 days to 1 day of simulated activity, and the time for assessing each candidate has been dropped from 20 hours to 8. Factor analyses of the 104 descriptors using data collected from assessment centers were analyzed several ways to verify that no unusual factor patterns or difficulties were found with field-based data. The study was constrained because only 113 data sets were available, when more than 312 would be required for complete analysis. However, the number of data sets was sufficient to divide the data into three different data sets based on different simulations the participants completed. Thirteen individual factor analyses across 3 different simulations found, with 1 exception within the 39 analyses, unidimensional factors. The six Management and seven Leadership constructs retained their construct integrity regardless of the way the factor analyses were structured, but some of the items forming more than one construct tended to load more heavily on a single factor because multiple skills were assessed within a given simulation. Data suggest that assessors tended to score all skills alike if the skills were assessed within the same simulation. Constructs retained their individuality if they were assessed independently. Results of this study verify with field data a constructually sound evaluation instrument. High reliability was found, based on individual skills, or when the skills were grouped into Management and Leadership skill configurations. Four appendixes describe the "vision" construct, give construct definitions, list some simulated activities associated with the inventory, and contain the skill-simulation matrix. (Contains 13 references.) (SLD)

ED 420 682

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Donald Coleman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Establishing Construct Validity and Reliability for the NAESP Professional Development Inventory:

Simplifying Assessment Center Techniques

By
Donald G. Coleman, Ed.D. Professor
Education Administration
California State University, Fresno
Fresno, California 93740
e-mail: don_coleman@csufresno.edu
Phone 209-278-0316; FAX 209-278-0404

Ron Areglado, Associate Executive Director
National Association of Elementary School Principals
1615 Duke Street
Alexandria, VA 22314
Phone 1-800-386-2377

and

R. C. Adams, Ph. D. Professor
Mass Communication and Journalism
California State University, Fresno
Fresno, California 93740

American Education Research Association
San Diego, California

April 1998

TM028357

Establishing Construct Validity and Reliability for the NAESP Professional Development Inventory:

Simplifying Assessment Center Techniques

This report documents the construct validity and reliability for the Professional Development Inventory (PDI), an assessment center sponsored by the National Association of Elementary School Principals (NAESP). The assessment center provides a means for assessing participants in situations simulating those confronted by principals on the job. The paper also explains attempts to reduce candidate time taken to be assessed as well as simplifying the scoring process to reduce assessor time essential for completing the assessing process. The NAESP assessment center was completely revised in 1997 changing from a two to a one day activity with new constructs and a simplified scoring system. The changes created an entirely new assessment configuration, easily implemented, but necessitating new validity and reliability studies on the reorganized instrument.

This paper not only explains the procedures used in revising and simplifying the assessment process, but examines the continuing research being done to assure validity and reliability. The most recent study is the third in a series already completed on the instrument, and documents data collected from the field on the first 113 participants undergoing assessment in four centers in three states using the revised format.

Background

The PDI represents a recent attempt to simplify assessment centers as a means for documenting on-the-job performance of professionals. Assessment centers have been used since the late 19th century to assess administrators but began in earnest in the U.S. with completion of a study at Harvard (Murray, 1938). The assessment process has been used extensively by the military and business and, more recently, in education administration (Hersey, 1977). The cost of using assessment centers, both in terms of time and energy, however, has been a limiting factor. The potential of assessment centers, therefore, remains latent as an assessment technique because of the complexities in design and cost-effectiveness in operation.

The PDI grew out of assessment work originally identified as the Administrative Diagnostic Inventory (ADI). Initial work was completed on the ADI in the mid-1980s at the University of Washington (Coleman & Hughes, 1993). NAESP purchased rights to the ADI in 1991; subsequent research, however, revealed a need to strengthen some measures and to modify the scoring system (Coleman, 1995; Jacobson, Parker, & Pinette, 1996; Sirotnik & Durden, 1996). At the time of its development, the ADI required candidates being assessed to complete two days of

simulated activities. The assessing process, moreover, required approximately 20 hours of assessor time for each candidate.

Given the criticisms aimed at both the scoring system and the excessive time required for assessing and scoring, NAESP officials authorized a reorganization of the assessing process to limit the time required to assess candidates to one day and to reduce the scoring time by half.

Using skills identified and recommended by various boards and professional organizations related to Education Administration, a team of practitioners, professors of Education Administration, and NAESP staff members redesigned the ADI, changing the simulation activities and devising measures for 13 skills determined to be the most critical among those suggested in the literature. As part of the re-design, the team purposely divided the skills selected to include both Management and Leadership in order to differentiate between essential personal and interpersonal administrative skills. The division resulted in the following classification scheme:

Management Skills

100 Planning
200 Organizing
300 Problem Solving
400 Creativity
500 Decisiveness
600 Systems Analysis

Leadership Skills

700 Vision
800 Communication
900 Instructional Leadership
1000 Group Leadership and Team Building
1100 Climate Development
1200 Moral Responsibility
1300 Instructional Analysis and Supervision.

Symbolic logic matrices (Langer, 1967) were used to create measures for the revised constructs (Creighton, Coleman, & Adams, 1997). Symbolic logic matrices provide a useful, but not perfect, tool to create discrete measures among constructs and for strengthening the verbal logic among descriptors when high inter-item correlations are required. Symbolic logic matrices, used here as truth tables, increase the statistical probability that desired outcomes can be achieved by checking the verbal logic among the variables at the point of creation. The statistical process verifying success in creating discrete constructs with high inter-item correlations is called "confirmatory" factor analysis. Confirmatory factoring procedures result when the constructs under development are known ahead of time, as was the case in this instance, and the items for measuring each construct needed to be created. To assure high inter-item correlation while achieving high construct integrity, a matrix was developed wherein the team members charged with reorganizing the instrument tested the verbal logic among items being created as they were developed. By creating "if-then" statements using each item created as an antecedent statement followed by using every other item in the construct as a "then" statement, the degree of compatibility between items could verify and illustrate graphically the logic of any combination among constructs.

After developing eight items for each of the 13 skill areas, an initial factor analysis on the instrument using a paper and pencil version of the instrument with 203 subjects produced limited success (Creighton, 1996). The initial study was an expeditious means to "confirm" the validity of the instrument created by the team of professionals. The data were factor analyzed using "R" factoring procedures with Kaiser normalization and varimax rotation. In "R" factor analyses, at least three highly inter-correlated items are deemed necessary to support construct identification. Five items with high inter-item correlations, of course, add assurance to correct construct identification but, also, add stability to both validity and reliability. To assure sufficiency in creating valid constructs, six statistical descriptors for each construct were desired for inclusion on the final version of the ADI. To assure a pool of items for selecting the best six items for each construct upon completion of the revision, eight descriptors were initially created for measuring each construct, as indicated in the sample in **Appendix A**.

As a result of the findings in this initial study, several construct definitions were rewritten (**see Appendix B**) for clarity and specificity, and items failing to correlate favorably within the constructs were revised. Additionally, in keeping with the NAESP philosophy that the instrument be created and used only for professional development rather than in administrator selection or evaluation, the name was changed at this point in the revision from the ADI to the PDI.

A repeated study with 146 teachers was next undertaken, again to confirm the construct validity using data collected from a second paper and pencil instrument. The paper and pencil study resulted in confirming 13 discrete constructs with eight highly inter-correlated measures (Coleman & Adams, 1997). The data collected from the revised pencil and paper survey, therefore, were used to establish a scoring system to score the instrument until enough participants had attended assessment centers to provide sufficient information to evaluate the scoring system with field-based data.

Purpose of the Study

The study reported herein, therefore, became the third study in the sequence. This study was designed to check both the practical operational aspects of the revisions as well as again verify the construct validity of the PDI using field-based data, or data collected in assessment centers. The statistical portion of the study, therefore, was designed to affirm that the constructs verified earlier using paper and pencil surveys could be replicated in the field as early as possible with actual data to justify continued use in assessments without placing additional participants at risk because of a faulty scoring system. Likewise, reliability of the instrument could also be determined.

Methodology

Since the instrument consisted of eight items in each of 13 skill areas, a total of 104 items, the number of subjects had to exceed 104 before the 104 by 104 correlation matrix could reach

closure using "R" factor analysis. Since implementing the revised PDI in the field, 113 candidates have been assessed, providing the earliest data pool for an adequate preliminary examination . This pool of data also provided the first opportunity to examine reliability of the instrument.

Instrumentation

All candidates completed a series of 12 activities (**see Appendix C**) that simulate those encountered by principals on a typical day at school. Several of the exercises require candidates to prepare written responses within the simulations while others require interactions with role players or other candidates that are recorded on video tape. The artifacts produced are later evaluated by a group of 24 assessors using the eight descriptors developed for each of the 13 skill areas. Additionally, each set of descriptors related to the skills is assessed at three different points in the assessment process by a separate team composed of two assessors each time **(see Appendix D)**. This means that all candidates undergoing assessment are evaluated on the 13 skills in differing types of simulated activities by 12 teams of assessors.

As can be seen in Appendix D, each team of two assessors assesses all candidates on a single simulation with as few as one skill to as many as five different skills being assessed by a team. Each assessor is trained in the assessment process and initially assesses each candidate independently. The two assessors assigned to assess scores related to a particular simulation next compare their individual scores to determine where differences in judgments exist in their independent thinking. Where differences exist, the two assessors must, then, reconcile their scores by re-examining the data and negotiating a settlement. Where irreconcilable differences in judgments exist, the scoring process requires the assessors to give candidates the benefit of the doubt. Inter-rater reliability, owing to no variance, is 1.00.

The instrumentation, therefore, consisted of 13 skills measured by a team of two different assessors evaluating each candidate at three different points among the simulations on the eight items related to the individual skills. Each item asks assessors to evaluate the candidate on a three point scale indicating Exceptional, Adequate, or Inadequate performance. The reconciled scores achieved by each team of assessors are entered into a computerized scoring system for calculating scores and printing individual profiles by NAESP.

Once the report containing information for each individual is completed, each candidate receives a comprehensive document profiling scores achieved on the skills assessed within each simulation as well as composite mean scores achieved on each skill across the three different activities. The report also includes mean scores for individual skills achieved on the three different simulations. Additionally, the printout also provides mean scores achieved on each descriptor on the three activities as scored by the three teams of assessors.

Since 104 variables are assessed at three points in each assessment, a total of 312 variables are used to develop the profile for each individual. However, to collect that many individual sets

of data requires an extensive period of time. In order to verify the data earlier, the design called for validating each of the three data sets of 104 variables individually. Therefore, all that was required for a preliminary analysis was to have an N exceeding 104. NAESP has collected, as indicated, 113 sets of information.

Population

The data collected came from assessment centers located in California, Georgia, and Texas. The candidates ranged from beginning administrators in entry level positions to experienced administrators with several years experience. Since the purpose of the study was to document the construct validity rather than to establish norms, experience was not considered a factor.

Data Collection and Analysis

Each participant arrives at a site by 8:00 a.m. After a 15 minute orientation, each candidate follows a unique schedule in attending the various activities. Some move to rooms where role players interact with them on various activities while others attend activities requiring only written responses. Lunch is frequently eaten while completing various activities, as done by many administrators, and the day ends at approximately 6:00 p.m. when the last assignment is completed.

The material completed is collected by the assessment center director and sorted by number rather than name of the participant to protect anonymity. The data are then ready for assessors to begin the assessing process, normally begun the next day.

Assessing the 113 candidates was completed slightly differently at each site. At a limited number of sites, the scoring was provided entirely by NAESP assessors while, at other sites, assessors were trained by NAESP staff before beginning the process. As assessments are repeated at the sites, experienced assessors become available to be assigned to the same activity and to train new assessors.

As indicated, each site enters the data into a computer program at the conclusion of the assessment and the data disks are sent to NAESP for scoring. These NAESP data were used in this study.

Since each candidate was assessed at three points during the assessment and each data set required an N greater than 104, the three data sets were necessarily considered to be separate measures. As indicated previously, to include all data in one large data set, an N greater than 312 would have been necessary. Waiting to collect more than 312 sets of information will take a considerable amount of time and could place an inordinate number of future candidates at risk if the scoring system proves inadequate. The decision was made to study continuously the results being achieved to ensure credibility. The following research procedures were implemented:

1) Anecdotal data were collected on participant satisfaction with the process and assessment center director observations regarding the time required to complete the analysis.

2) Frequency distributions were calculated for each item and screened for normality in skewnesss and kurtosis.

3) Each of the 13 sets of eight items related to each of the separate skills was subjected to a factor analysis to determine if all items created for a particular dimension and assessed in a unique simulation loaded on a single factor. Therefore, 39 factor analyses were conducted to determine if the construct retained its integrity when used in the various activities.

4) The Management and Leadership constructs from the three data sets were separated and subjected to factor analyses to evaluate how Management items loaded across the six Management dimensions and how Leadership items loaded across the seven Leadership dimensions.

5) Data from all 13 dimensions in the three data sets were factor analyzed to determine the factor structure when all 104 variables were included. A minimum factor loading of .50 was set as the desired limit in determining adequate loadings. All 13 data sets were combined for each of the three different simulations to determine if the three independent measures retained independent constructs or formed a single set of constructs.

6) The final set of 13 factor analyses consisted of grouping the three simulation data sets collected on each skill and determining if the data form a unidimensinal factor or multidimensional factors, across simulations.

7) Reliability coefficients were calculated for each of the 13 scales and for the relevant combinations---i.e., Management and Leadership.

Findings

For brevity, only a narrative describing the findings is provided. (Detailed tables on the data are available from NAESP or the authors.) The statistical analyses, however, led to the following findings:

1) Analysis of the One Day Assessment and Assessing Time

From anecdotal feedback from directors and participants, the transition from a two to one day activity was favorably received. Comments from directors with experience with both the former and revised assessment centers were very favorable. Participants indicated that the simulations were authentic, when compared to a normal school day, but having cam-corders present in interactions was not a normal situation. Most participants, however, believed that, as they became absorbed in the various activities, the cam-corder was not a distractor.

The time for assessing was also reduced by half, to 10 hours for each participant, when 12 candidates were assessed; but at one site, the time dropped to eight hours for each participant when

eight candidates were assessed by experienced assessors. Experienced assessors with fewer candidates, apparently, require considerably less time to assess.

2) Descriptive Data Results

The descriptive data collected on all three data sets revealed but few abnormalities in the distributions. For the factor analyses, these were addressed by distributional transformations (Rummel, 1970) to assure a good approximation to a bivariate, normal distribution when estimating the correlation coefficients.

3) Analysis of 13 Constructs

To determine if each of the constructs maintained construct integrity, the three independent assessments were identified by Simulations as A, B, and C. As described above, this means that data from each of the three assessments came from information collected from data sets grouped by the different simulated activities **(see Appendix E).** A factor analysis, therefore, was conducted for simulations A, B, and C separately for each of the skills. Since the data represented 13 skills, 39 factor analyses were calculated.

The individual factor analyses of the eight descriptors developed for each of the 13 dimensions revealed a single, unidimensional construct within each of the 39 data sets with one exception. When Vision was assessed within the simulation, Vision Statement, the descriptors separated into two factors. This division did not appear on the original documentation or on the other two comparable factor analyses. The first four items loaded on one construct and the last four on a second. The reader can find the list of descriptors developed for Vision in Appendix A. Each of the other constructs, however, possessed a consistent cluster of items with high inter-item correlations irrespective of the simulations to which they were applied.

4) Analysis of Six Management Constructs

When the Management items were separated into Simulations A, B, and C and subjected to factor analyses, the six constructs did not maintain complete construct integrity as was the case in the former study.

Simulation A: The six constructs factor analyzed in simulation A, for instance, resulted in five factors with eigenvalues greater than 1.00 and accounting for 76.5% of the variance. The Planning and Organizing skills loaded on Factor One while Problem Solving and Decisiveness loaded on Factor Two. The other two skill areas, Creativity and System Analysis, retained their separate identities. A fifth factor did not produce items with factor loadings at the .50 level, but were indicative of the last four items on the Decisiveness skill.

Simulation B: The six constructs factor analyzed in Simulation B resulted in four factors with eigenvalues greater than 1.00 accounting for 68.3% of the variance. Problem Solving and Decisiveness loaded on the first factor, Creativity and Systems Analysis loaded on the second, Planning loaded on the third, and Organizing loaded on the fourth factor.

Simulation C: The six constructs factor analyzed in Simulation C resulted in six factors with eigenvalues greater than 1.00 and accounting for 69.1% of the variance. Organization and Problem Solving loaded on the first, Systems Analysis on the second, Creativity on the third, Planning on the fourth, and Decisiveness on the fifth factor.

Analysis of Seven Leadership Constructs

When the Leadership items were separated for Simulations A, B, and C and subjected to factor analyses, the seven constructs did not maintain complete construct integrity as was the case in the former study.

Simulation A: The seven constructs factor analyzed in Simulation A resulted in six factors with eigenvalues greater than 1.00 and accounting for 65.4% of the variance. Communications and Instructional Leadership loaded on the first factor while the others loaded on separate factors identifying Climate Development, Vision, Group Leadership and Team Building, Instructional Leadership and Supervision, and Moral Responsibility, in order.

Simulation B: The seven constructs factor analyzed for Simulation B resulted in seven factors with eigenvalues greater than 1.00 and accounting for 66.4% of the variance. Communications and Instructional Analysis and Supervision loaded on the first factor while Group Leadership and Team Building and Vision loaded on the second followed by Climate Development on Factor three. Factor four consisted of Instructional Leadership and the last four items from Vision and a weak representation from Moral Responsibility. Factor six included items from Instructional Analysis and Supervision for a second time in this section and factor seven was ill defined.

Simulation C: The seven constructs factor analyzed in Simulation C resulted in six factors with eigenvalues greater than 1.00 and accounting for 66.5% of the variance. Vision, Instructional Leadership, Group Leadership, and Moral Responsibility loaded on the first factor; Communications and Instructional Analysis and Supervision loaded on the second factor; and Climate Development loaded on the third. Group Leadership, Instructional Leadership, and Instructional Analysis and Supervision loaded separately on factors four, five, and six. As can be seen above, the latter three appeared in the factor matrix for a second time but were weak constructs with factor loadings of less than .50.

5) Analysis of Management and Leadership Skills Combined

When the Management and Leadership items were combined for Simulations A, B, and C and subjected to factor analyses, the 13 constructs did not maintain complete construct integrity as was the case in the former study.

Simulation A: The 13 skill constructs factor analyzed for Simulation A resulted in 14 factors with eigenvalues greater than 1.00 and accounting for 72% of the variance. Problem Solving, Decisiveness, and Vision loaded on the first factor; Planning and Organizing loaded on the second factor; Communications and Instructional Leadership loaded on the third factor; Climate

Development loaded on the fourth; Creativity loaded on the fifth; Systems Analysis and Design loaded on the sixth; Group Leadership and Team Building loaded on the seventh; Instructional Analysis and Supervision loaded on the eighth; and Moral Responsibility loaded on the ninth factor. The remaining factors had so few significant loadings as to make the construct definition indistinguishable.

Simulation B: The 13 skill constructs factor analyzed for Simulation B resulted in 12 factors with eigenvalues greater than 1.00 and accounting for 70.6% of the variance. The 13 constructs loaded clearly on the first seven factors. Group Leadership and Team Building, Vision, Moral Responsibility, and Instructional Leadership loaded on Factor One; Problem-Solving and Decisiveness both loaded on Factor Two; Creativity and Systems Analysis and Design loaded on Factor Three; Communications and Instructional Analysis and Supervision loaded on Factor Four; Planning, Climate Development, and Organizing loaded on factors five through seven, respectively. Factor Eight provided slight indication that the last four items in Instructional Leadership were forming a weak but separate factor; Factor Nine revealed a construct measuring Instructional Analysis and Supervision for a second time, but with much lower factor loadings; in a separate but weak factor, Factor Ten picked up the last four items in Moral Responsibility. The balance of the factors were indistinguishable.

Simulation C: The 13 skill constructs factor analyzed for Simulation C resulted in 13 factors with eigenvalues greater than 1.00 and accounting for 72.4% of the variance. All 13 skills loaded on the first six Factors with the balance of the factors picking up identification of selected skills very weakly a second time. Creativity, Vision, Instructional Leadership, Group Leadership and Team Building, and Moral Responsibility loaded on Factor One; Factor Two included Organization, Problem-Solving, and Climate Development; factor three included Communications and Instructional Analysis and Supervision; factor four Systems Analysis and Design; Factor Five Planning; and Factor Six Decisiveness.

6. Thirteen Factor Analyses on the each of the Skills Combining Data from the Three Observations

Data from the three observations for each of the 13 skills were combined into a single file and factor analyzed to determine if the data would form a unidimensional factor or multi-dimensional factors. The 13 analyses revealed that, with one exception, the three observations revealed three factors, for the three simulations, with eigenvalues greater than 1.00. The exception was Moral Responsibility, which formed a weak fourth factor.

7. Reliability

Alpha reliability coefficients (Cronbach, 1951) were calculated on the 13 skill areas and the three combined Management and Leadership skills as follows:

Skill /Skill Area	Alpha
Planning	.9448
Organizing	.9410
Problem-Solving	.9304
Creativity	.9646
Decisiveness	.8863
Systems Analysis and Design	.9467
Vision	.9518
Communications	.9159
Instructional Leadership	.8917
Group Leadership and Team Building	.9508
Climate Development	.9681
Moral Responsibility	.9162
Instructional Analysis and Supervision	.9225

Group	Skill	Skill Area	
A	1-6	Management	.9156
B	1-6	Management	.9245
C	1-6	Management	.9299
A	7-13	Leadership	.9231
B	7-13	Leadership	.9609
C	7-13	Leadership	.9449

In sum, the 13 eight-item scales provide from 78% to 94% reliable variance for the individual skill evaluations, Management and Leadership combinations from 84% to 92% reliable variance in evaluating these clusters over three simulations, individually. The former represent highly reliable measurement over three simulations, in the hands of multiple raters. The latter represents high reliability over subscales, in the hands of multiple raters.

The Alpha coefficient in Simulation A was estimated at .939; for B, .956, and C, .962, while the overall instrument Alpha coefficient was estimated at .978

Conclusions

The findings, based upon these preliminary data, lead to the following conclusions:

1) From anecdotal information collected from both assessment center directors and participants, the conversion from a two to one day activity has been successful. Not only has the time taken to assess the candidates been reduced from two days to one, but the scoring time to assess each candidate has been cut to 40 percent with skilled assessors assessing eight candidates rather than 12 at a time. Further reduction in time can be expected when the number of descriptors for each

skill is reduced from eight to six, but this reduction is not expected to net the huge time saving achieved by the other modifications. The participants also reported informally that the assessment is authentic and nearly all gave the experience high praise.

2) The descriptive data appeared to be normal except in a limited number of cases. Few corrections had to be made to approximate normal distribution patterns among responses. The descriptive data, therefore, looked very good.

3) When the data were divided into the three independent assessments across the various simulations, identified as simulations A, B, and C, the data revealed that the skill constructs retained their discrete identities or construct integrity. The skill, Vision, separated into two factors only when linked to the simulated activity associated with "Vision Statement." In earlier construct analysis studies as well as on the other two analyses associated with the current study, this division in the construct failed to appear. Likewise, the division did not appear in any of the other analyses conducted in the investigation discussed here. Therefore, the final conclusion must wait until sufficient data are collected (in excess of 312 data sets).

Having stated this cautionary comment, however, Decisiveness and Moral Responsibility skills, in addition to Vision, in at least one of the analyses, gave slight indications, but not significant, that the last half of the items on each skill may be providing factors separate from the first four items. However, no conclusion is reached here except to point out that the phenomena need to be re-examined when sufficient data are collected to complete the studies.

4) The Management skill constructs maintained item integrity when subjected to all three factor analyses within each differing set of simulated activities, i.e., A, B, C. However, the first and second factors possessed a pattern of multiple constructs loading with one another on the factor. The multiple loadings, however, were inconsistent among the three analyses. Nonetheless, each Management construct maintained a separate identity in at least one of the three simulations with the exception of Problem Solving. Interestingly, this same phenomenon occurred in the second paper and pencil research study. In that study, Problem Solving loaded nearly equally on three of the other factors: Planning, Organizing, and Creativity. In the earlier study, Problem Solving retained its own identity when all 13 Management and Leadership factors were analyzed together.

Additionally, examination revealed that the multiple loadings of constructs on common factors apparently are a result of specific simulations calling for a single team of assessors to provide assessments of multiple skills. These teams of assessors, knowingly or unknowingly, evaluate a candidate in a similar pattern across all items when multiple skills are assessed. That is , the assessors appear to get a particular mind-set of the candidate on the first skill area and assess the candidate similarly across all of the other skills being assessed in the same simulation. The effect was to cause each simulation to retain its separate identity or integrity as well as the items within a skill still possessing high item inter-correlations.

This phenomenon occurred in all three factor analyses indicating that the simulations were an over-riding variable within the factor analyses. Obviously, the factor patterns changed depending upon the simulation because the skills related to the factors changed for each factor analysis. However, when the variance resulting within each of the three simulation patterns was examined, the number of skills assessed in a particular simulation was a dominant force. Indications are that both skills and simulations appear as maintaining a degree of integrity in the factoring process.

What remains unknown is what will happen to the skill patterns when sufficient data (N > 312) are gathered to calculate a factor analysis on all 312 items simultaneously. The assumption is that the patterns will become more stable because all 13 constructs will be subjected to a common factor analysis rather than partialing data as was done in this study.

The Leadership skill constructs followed a pattern similar to the Management skills. The inter-item correlations retained good construct integrity, but several skills loaded on the first and second constructs following the pattern found in the simulations. All of the Skills with the exception of communications, however, retained a separate identity at least once in the three different simulation combinations. Unlike the paper and pencil survey, Communications skills in all three instances were linked to another skill or combination of skills. Where a particular set of skills required a team of assessors to assess multiple skills, the team tended to evaluate the candidates on all skills alike. Again, the three skill assessments (A, B, or C) loaded differently, depending on the skills required to assess in particular simulations. Since all three skill assessments could not be assessed simultaneously, final conclusions rest on the assumption that a clearer pattern distinction will emerge when sufficient data are collected.

5) The Management and Leadership skills constructs when combined formed a pattern much like those found in the Management and Leadership constructs when calculated separately. The individual items retain their cluster identity very well, but the items form multiple clusters on the factors, depending upon the simulations being assessed. That is, the more skills required to be assessed for a particular simulation, the greater number of skills that clustered on a single factor. The simulation with the greatest number of skills loaded on Factor One; the simulation with the second most factors assessed loaded on Factor Two and so on until all skill clusters were accounted for. Presumably, this unusual clustering configuration will disappear when all 312 items are factor analyzed with a sufficient number of participants (N >312).

As was the case in the Management and Leadership analyses, 10 of the 13 skills retained their individual identities in at least one of the three sets of simulated activities. The only exceptions included Communications, Instructional leadership, and Vision, which loaded on a factor in combination with one or more other skills.

6) When each of the 13 skills were factor analyzed with all three simulations together, the clusters retained their individual identities. This indicates that the three sets of simulations are discrete or independent of each other. What will occur when the simulations are combined in a single analysis with a number of participants greater than 312 is not clear. Again, the data may or may not form unidimensional constructs.

7) The data reflect high reliability coefficients for individual skills as well as for the Management and Leadership item clusters. Final instrument reliability will depend upon data from a data set of participants with an N greater than 312.

8) The three step process described here to develop the revised PDI was simplified by the three stage process: two paper and pencil surveys to document the constructs and one with field-based data to verify the paper and pencil findings. The constructs thus developed provide a valid and reliable instrument when subjected to field testing.

Summary

NAESP has developed a one day assessment center to assess Education Administrators on 13 skills. The time for assessing candidates has been reduced from two days to one and the time for assessing each candidate dropped from 20 hours for each candidate to eight while retaining high regard for the process from participants.

Factor analyses of the 104 descriptors using data collected from assessment centers were factor analyzed several ways to verify that no unusual factor patterns or difficulties were found within field-based data. The study was constrained primarily because only 113 data sets were available when more than 312 were required for a complete analysis. The number of subjects is sufficient, however, to divide the data into three different data sets based upon different simulations that the participants completed. The study was a check to discover as early as possible any inadequacies that could be found within the assessment structure. Thirteen individual factor analyses conducted across three differing simulations found, with one exception within the 39 analyses, unidimensional factors.

The six Management and seven Leadership constructs retained construct integrity irrespective of to how the factor analyses were structured. However, some of the items forming more than one construct tended to load more heavily on a single factor because of multiple skills being assessed within a given simulation. The data suggest that assessors tended to score all skills alike if the skills were assessed within the same simulation. This phenomenon occurred irrespective of the various simulations assessed. Constructs retained their individuality if they were assessed independently, but simulations retained their identity as well when multiple skills were assessed giving a "situation specific" indication.

The results of the study verify with field data a constructurally solid evaluation instrument. Data collection must continue until an N greater than 312 is achieved so all data, rather than partial, can be calculated simultaneously.

The instrument was found to possess high reliability, based upon individual skills and when the skills were grouped into Management and Leadership skill configurations.

References

Coleman, Donald G. (November, 1995) *Construct Validity: Administrative Diagnostic Inventory.* Fresno: California State University, Fresno. **Report for National Association of Elementary School Principals.**

Coleman, Donald G. (November, 1996) *Construct Validity of the Revisions of Administrative Diagnostic Inventory.* Fresno: California State University, Fresno. **Report for the National Association of Elementary School Principals.**

Coleman, Donald G. and Adams, R.C. (August, 1997) *Construct Validation of the Professional Development Inventory: A Report on Organizing the NAESP Assessment Center.* Vail: **National Council for Professors of Education Administration**.

Coleman, Donald G. and Hughes, Larry W. (1993) *History and Philosophy of the Administrative Diagnostic Inventory: An Interview with Dale Bolton.* Alexandria, VA: **National Association of Elementary School Principals**.

Creighton, Theodore B. (1996) *A Construct Validation of the Administrative Diagnostic Inventory.* (Doctoral Dissertation). California State University; University of California, Davis.

Creighton, Ted, Coleman, Donald G., and Adams, R.C. (1997, September) *A Construct Validation of the Administrative Diagnostic Inventory.* Washington D. C.: **Education Clearinghouse on Assessment and Evaluation** (TM 027697).

Cronbach, L. J. (1951) *Coefficient Alpha and the Internal Structure of Tests.* **Psychometrika,** *16, 291-335.*

Hersey, P. W. (1977) *NAACP's Assessment Center--from Concept to practice."* **NASSP Bulletin.** 61:410. September.

Jacobson, Larry; Parker, Stephanie; and Pinette, Susan. (1996) *360 0 Assessment: Capturing the Performance of Instructional Leaders Using Simulated and Situated Exercises.* Washington D.C.: **International Congress on the Assessment Center** Method.

Langer, Susanne K. (1967) **An Introduction to Symbolic Logic** *(3rd Ed. Rev.)* New York: Dover Publications, Inc.

Murray, H. A. (1938) **Explorations in Personality.** New York: Oxford University Press.

Rummel, R. J., (1970) **Applied Factor Analysis**. Evanston: Northwestern University Press.

Sirotnik, Kenneth A. and Durden, Phyllis. (October, 1996) *The Validity of Administrator Performance Systems : The ADI as a case in point.* **Educational Administration Quarterly.** Vol. 32, No. 4.

Appendix A

0700	**VISION** Possesses a clear and positive view of the past, present and future of the school..

0701	holds a clear, positive view of the school.
0702	possesses a vision of an ideal school
0703	interprets the vision to the school community
0704	seeks a shared vision of the school
0705	sets programs and procedures in the context of the school's vision
0706	encourages activities supporting the school vision
0707	nurtures the vision of the school
0708	maintains the school vision in daily activities

Appendix B

PDI Construct Definitions

0100 PLANNING	Defines purpose and sets organizational goals.
0200 ORGANIZING	Defines tasks to be completed in various activities and sequences events
0300 PROBLEM SOLVING	Analyzes problems effectively and reaches logical conclusions.
0400 CREATIVITY	Demonstrates innovation and inventiveness in work related situations.
0500 DECISIVENESS	Renders timely and appropriate decisions.
0600 SYSTEMS ANALYSIS and DESIGN	Integrates various organizational components into a coherent, and effective operational pattern.
0700 VISION	Possesses a clear and positive view of the past, present and future of the school.
0800 COMMUNICATIONS	Processes messages with precise understanding
0900 INSTRUCTIONAL LEADERSHIP	Works effectively with the school community to enhance student learning.
1000 GROUP LEADERSHIP & TEAM BUILDING	Mobilizes others to collaborate in accomplishing school goals and solving problems.
1100 CLIMATE DEVELOPMENT	Shapes the psycho-social environment of the school to promote accomplishment of the mission.
1200 MORAL RESPONSIBILITY	Demonstrates universally held core values and beliefs.
1300 INSTRUCTIONAL ANALYSIS AND SUPERVISION	Works effectively with teachers to improve instruction.

Appendix C

Simulated Activities Associated with the Professional Development Inventory

Each candidate in the assessment center completes activities as directed associated with the following simulations:

1) **Vision Statement**

Instructional Analysis

2) **Lesson Analysis**
3) **Plan Conference**
4) **Conduct Conference**
5) **Written summary**

6) **Case: Persistent Parent**

Encounters

7) **Teacher**
8) **Students**
9) **Teacher and Teacher Aide**

10) **Candidate Selection**

11) **In-basket**

Priority
 # 1
 # 6
 # 7
#10

12) **Group Discussion**

Appendix D
Skill-Simulation Matrix
for NAESP Assessment Center

Simulations / Skills	Planning	Organizing	Problem Solving	Creativity	Decisiveness	Systems Analysis	Vision	Communications	Instructional Leadership	Group Leadership & Team Building	Climate Development	Moral Responsibility	Instructional Analysis & Supervision
Vision Statement							X		X	X		X	
IA: Lesson Analysis													X
IA: Plan Conference	X												
IA: Conduct Conference								X				X	X
IA: Written Summary								X					X
Case: Persistent Parent		X	X								X		
Enc: Teacher				X	X	X		X	X				
Enc: Student			X		X						X		
Enc: Teacher/Aide						X				X	X		
Selection			X		X		X						
IB: Priority	X	X											
IB: #1						X							
IB: #6				X									
IB: #7		X											
IB: #10	X												
Group Discussion				X			X		X	X		X	

Code:
IA =Instructional Analysis
Enc=Encounter
IB =In-basket

Appendix E
Skill-Simulation Matrix by the Three Independent Assessments on the Professional Development Inventory

Skill	Simulation Configuration A	B	C
1) Planning	IB Priority	IB#10	IA: Plan Conference
2) Organizing	IB Priority	IB#7	Persistent Parent
3) Problem-Solving	Selection	Enc: Student	Persistent Parent
4) Creativity	IB#6	Enc: Teacher	Group Process
5) Decisiveness	Selection	Enc: Student	Enc: Teacher
6) Systems Analysis and Design	IB#1	Enc: Teacher	Enc: Teach/Aide
7) Vision	Selection	Vision Statement	Group Process
8) Communications	Enc: Teacher	IA: Conduct Conference	IA: Written Summary
9) Instructional Leadership	Enc: Teacher	Vision Statement	Group Process
10) Group Leadership and Team Building	Enc: Teach/Aide	Vision Statement	Group Process
11) Climate Development	Enc: Student	Enc: Teach/Aide	Persistent Parent
12) Moral Responsibility	IA: Conduct Conference	Vision Statement	Group Process
13) Instructional Analysis and Supervision	IA: Instructional Analysis & Suprv	IA: Conduct Conference	IA: Written Summary

Code:

IB: In-basket
Enc: Encounter
IA: Instructional Analysis